DOCUMENT RESUME

ED 222 508 TM 820 651

AUTHOR Stocking, Martha L.; Lord, Frederic M.
TITLE Developing a Common Metric in Item Response Theory.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
REPORT NO ETS-RR-82-25-ONR
PUB DATE Jun 82
CONTRACT N00014-80-C-0402
NOTE 44p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Data Analysis; *Equated Scores; Item Analysis; Item Banks; *Latent Trait Theory; *Scaling; Secondary Education; Test Construction; Testing Problems; Test Items
IDENTIFIERS *Item Calibration

ABSTRACT

A common problem arises in scale transformation when independent estimates of item parameters from two separate data sets must be expressed in the same metric. These item parameter estimates will be different because the metric or scale defined by each independent calibration of the items is different. The problem is frequently confronted in studies of horizontal and vertical equating and in studies of item bias. The nature of scale transformation from one metric to another is discussed, and the use of mean and sigma method is reviewed. A new method is presented which uses more information from item calibrations to yield an estimated item response function or item characteristic curve. Using data from 12 separate administrations of the Scholastic Aptitude Test, the mean and sigma transformation and characteristic curve method are shown to work well in the verbal data and most of the mathematical data. When the mean and sigma method was unsatisfactory, the information content of the characteristic curve method worked better. In particular, the characteristic curve method produced a much better transformation for the item discriminations. (CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED222508

RR-82-25-ONR

DEVELOPING A COMMON METRIC IN ITEM RESPONSE THEORY

Martha L. Stocking
and
Frederic M. Lord

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Office of Naval Research

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

This research was sponsored in part by the Personnel and Training Research Programs Psychological Sciences Division Office of Naval Research, under Contract No. N00014-80-C-0402

Contract Authority Identification Number NR No. 150-453

Frederic M. Lord, Principal Investigator

Educational Testing Service
Princeton, New Jersey

June 1982

Reproduction in whole or in part is permitted for any purpose of the United States Government.

Approved for public release; distribution unlimited.

TM 820 651

DEVELOPING A COMMON METRIC IN ITEM RESPONSE THEORY

Martha L. Stocking

and

Frederic M. Lord

This research was sponsored in part by the Personnel and Training Research Programs Psychological Sciences Division Office of Naval Research, under Contract No. N00014-80-C-0402

Contract Authority Identification Number NR No. 150-453

Frederic M. Lord, Principal Investigator

Educational Testing Service

Princeton, New Jersey

June 1982

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Developing a Common Metric in Item Response Theory		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER RR-82-25-ONR
7. AUTHOR(s) Martha L. Stocking and Frederic M. Lord		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0402
9. PERFORMING ORGANIZATION NAME AND ADDRESS Educational Testing Service Princeton, NJ 08541		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 150-453
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research (Code 458) Arlington, VA 22217		12. REPORT DATE June 1982 13. NUMBER OF PAGES 29
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Item Response Theory, Common Metric, Scale Transformations, Item Banking, Equating, Item Bias

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A common problem arises when independent estimates of item parameters from two separate data sets must be expressed in the same metric. This problem is frequently confronted in studies of horizontal and vertical equating and in studies of item bias. This paper discusses a number of methods for transforming one metric to another metric and presents a new method. Data are given comparing this new method with a current method and recommendations are made.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Developing a Common Metric in Item Response Theory

Abstract

A common problem arises when independent estimates of item parameters from two separate data sets must be expressed in the same metric. This problem is frequently confronted in studies of horizontal and vertical equating and in studies of item bias. This paper discusses a number of methods for transforming one metric to another metric and presents a new method. Data are given comparing this new method with a current method and recommendations are made.

Developing a Common Metric in Item Response Theory*

Introduction

Suppose that item parameters for a given set of items have been independently estimated using data obtained from two different groups of examinees. These item parameter estimates will be different because the metric or scale defined by each independent calibration of the items is different. Many applications of item response theory (IRT) require that these item parameter estimates be expressed in the same metric. Such applications include vertical score-scale equating, horizontal score-scale equating, and item bias studies.

It is possible to transform item parameter estimates in one metric to another metric by a number of different methods. This paper will discuss the nature of these scale transformations, survey a number of current transformation methods, and present a new method and some results of its application.

The Nature of Scale Transformations

Item response theory models $P_i(\theta_a;\alpha_i,\beta_i,\gamma_i)$, the probability of a correct response to item i by a person with ability level θ_a. In typical models, $P_i(\theta_a;\alpha_i,\beta_i,\gamma_i)$ is a function of $\alpha_i(\theta_a - \beta_i)$, where α_i is the item discrimination, β_i is the item difficulty,

---

*This work was supported in part by contract N00014-80-C-0402, project designation NR150-453 between the Office of Naval Research and Educational Testing Service. Reproduction in whole or in part is permitted for any purpose of the United States Government.

and γ_i is the probability that an individual of very low ability answers the item correctly. When $P_i(\theta_a;\alpha_i,\beta_i,\gamma_i)$ is a function of $\alpha_i(\theta_a - \beta_i)$, the origin and unit of measurement of the ability (and difficulty) metric are undetermined. That is to say, suppose θ_a is transformed by a linear transformation, producing θ_a^*. Suppose the same linear transformation is applied to β_i to produce β_i^*. Finally, α_i is divided by the multiplicative constant of the linear transformation to produce α_i^*. These transformations will not change the probability of a correct response: $P_i(\theta_a^*;\alpha_i^*,\beta_i^*,\gamma_i) = P_i(\theta_a;\alpha_i,\beta_i,\gamma_i)$. Notice that no transformation is necessary for the γ_i because γ_i is on the probability metric.

If an item is calibrated, i.e., its parameters are estimated, as part of one test, and then calibrated as part of a second test given to a different group, the actual values of the estimates of the parameters will differ because the scales established by the two calibrations differ. However, the relationship between these two scales will be linear since they differ only in origin and unit of measurement.

If b_{i1} is the estimate of item difficulty from the calibration of item i in test 1, and b_{i2} is the estimate of the same item difficulty from the calibration of test 2, b_{i2}^*, the value of b_{i2} transformed to the scale of test 1, is

$$b_{i2}^* = Ab_{i2} + B \quad , \tag{1}$$

where A and B are constants of the linear transformation of scale. If estimated item difficulties are transformed by a linear transformation, estimated abilities must be transformed by the same transformation, thus

$$\hat{\theta}^{*}_{a2} = A\hat{\theta}_{a2} + B \quad . \tag{2}$$

If estimated item difficulty and ability are transformed by these linear expressions, then estimated item discrimination is transformed by

$$a^{*}_{i2} = a_{i2}/A \quad . \tag{3}$$

These transformations do not change $a_{i2}(\hat{\theta}_{a2} - b_{i2})$, consequently $P_i(\hat{\theta}_{a2};a_{i2},b_{i2},c_{i2}) = P_i(\hat{\theta}^{*}_{a2},a^{*}_{i2},b^{*}_{i2},c_{i2})$.

The problem of transforming the scales reduces to the problem of finding the appropriate A and B of the linear transformation. If we were dealing with true values of the parameters on their respective scales, it would be simple to find the correct values of A and B ; we could plot the values of two or more item difficulties and determine the line passing through them. But, we do not have true values; we have only estimates of them, and these estimates contain error. The estimated item difficulties will not fall into a straight line, but be scattered around some straight line. All methods of transforming scales attempt to estimate the parameters of this line by various techniques, and are applicable to any IRT model where $P_i(\theta_a;\alpha_i,\beta_i,\gamma_i)$ is a function of $\alpha_i(\theta_a - \beta_i)$.

Current Methods

Superficially, the problem of finding the linear relationship between two sets of numbers might seem to call for simple regression techniques. The estimated item difficulties (or abilities) from one calibration might be used as the independent variable, and those obtained from the second calibration as the dependent variable. This approach would be incorrect. A regression approach assumes the independent variable is measured without error; we know this is not the case. But more important, a regression procedure is not symmetric with respect to its treatment of the two estimates of item difficulties. Since we have no reason for emphasizing or favoring one estimate of item difficulty over another estimate of the same item difficulty, we require a symmetric procedure.

A class of symmetric methods uses the first two moments of the distributions of estimated item difficulties. These methods find the parameters of the linear transformation, A and B, such that the mean and standard deviation of the transformed distribution of estimated item difficulties from the second calibration are equal to the mean and standard deviation of the estimated item difficulties from the first calibration.

A simple application of this method is found in Marco (1977) and in Cook, Eignor, and Hutten (1979). Poorly estimated item difficulties may have a serious impact of the computation of sample moments, however, producing a linear transformation that cannot be useful. Cook et al. (1979)

attempt to solve this by restricting the range of the difficulties used in computing moments.

Bejar and Wingersky (1981) use a more elaborate approach. Robust methods that give smaller weights to outlying points are used to estimate the moments. Linn, Levine, Hastings, & Wardrop (1980) attempt to reduce the influence of outliers by using weighted moments where the weights are inversely proportional to the estimated standard error of the estimates of the item difficulties.

The Bejar and Wingersky procedure treats all outliers in the same fashion, regardless of their standard error. The Linn et al. procedure treats all points with the same standard error in the same fashion, regardless of their outlier status. A procedure was developed by Lord and Stocking which attempts to overcome these potential problems. This procedure begins with a weighted estimate of the transformation exactly as in Linn et al. A robust procedure is then used to give small weights to those values whose perpendicular distance from this initial line is large, and a new line is estimated. The robust weighting is repeated until changes in the perpendicular distances become small. Details of this method are presented in the Appendix. Some results of this method will be described in subsequent sections of this paper.

A drawback of all of these "mean and sigma" transformation procedures is that they are typically applied only to the estimated item difficulties. That is, the A and B of the linear transformation of scale are estimated using only the b_i, and then applied to transform the $\hat{\theta}_a$ and the a_i. While this is theoretically correct, better methods may exist which use more of the information available from the calibrations.

A class of methods, called "characteristic curve methods" in this paper, uses more information from calibrations. Each calibration of an item yields an estimated item response function or item characteristic curve $\hat{P}_i(\theta_a) \equiv P_i(\theta_a; a_i, b_i, c_i)$. If estimates were error free, the proper choice of A and B for the linear transformation would cause these two curves to coincide. Haebara (1980) averages the squared difference between the individual item response functions over a suitable distribution of θ , sums over the items common to the two calibrations, and chooses A and B to minimize this sum. Divgi (1980) chooses the A and B of the linear transformation to minimize the maximum difference between the sum of item response functions for the first calibration and the sum of the item response functions for the second calibration.

The New Method

This method falls into the class of characteristic curve methods. An examinee, a , with ability θ_a has a true score ξ_a defined by

$$\xi_a \equiv \xi(\theta_a) \equiv \sum_{i=1}^{n} P_i(\theta_a; \alpha_i, \beta_i, \gamma_i) \quad , \tag{4}$$

where n is the number of items in the test. The correct linear transformation of scales from two different calibrations of the same test would

produce the same true scores for examinee a if the α_i , β_i , γ_i were known. If $\hat{\xi}^*_a$ is the estimated true score obtained from the second calibration of the test after it has been transformed to the scale of the first, then

$$\hat{\xi}^*_a \equiv \hat{\xi}^*(\theta_a) \equiv \sum_{i=1}^{n} P_i(\theta_a; a^*_i, b^*_i, c_i) \quad . \tag{5}$$

For an examinee, the difference $(\hat{\xi}_a - \hat{\xi}^*_a)$ should be small. In practice, we want to choose A and B such that for a suitable group of examinees, the average squared difference between true score estimates is as small as possible. The function to be minimized is

$$F = \frac{1}{N} \sum_{a=1}^{N} (\hat{\xi}_a - \hat{\xi}^*_a)^2 \quad , \tag{6}$$

where N is the number of examinees in the arbitrary group.

This function F considered as a function of A and B will be minimized when

$$\frac{\partial F}{\partial A} = \frac{-2}{N} \sum_{a=1}^{N} (\hat{\xi}_a - \hat{\xi}^*_a) \frac{\partial \hat{\xi}^*_a}{\partial A} = 0 \quad , \tag{7}$$

and

$$\frac{\partial F}{\partial B} = \frac{-2}{N} \sum_{a=1}^{N} (\hat{\xi}_a - \hat{\xi}^*_a) \frac{\partial \hat{\xi}^*_a}{\partial B} = 0 \quad . \tag{8}$$

Now, using the chain rule of differentiation,

$$\frac{\partial \hat{\xi}^*_a}{\partial A} = \sum_{i=1}^{n} \left(\frac{\partial P_i(\theta_a; a^*_{i2}, b^*_{i2}, c_{i2})}{\partial b^*_{i2}} \frac{\partial b^*_{i2}}{\partial A} + \frac{\partial P_i(\theta_a; a^*_{i2}, b^*_{i2}, c_{i2})}{\partial a^*_{i2}} \frac{\partial a^*_{i2}}{\partial A} \right) . \quad (9)$$

Differentiating equations (1) and (3) gives $\frac{\partial b^*_{i2}}{\partial A} = A$ and $\frac{\partial a^*_{i2}}{\partial A} = \frac{-a_{i2}}{A^2}$.

Substituting these derivatives into (9) gives the partial derivative

$$\frac{\partial \hat{\xi}^*_a}{\partial A} \equiv \sum_{i=1}^{n} \left(b_{i2} \frac{\partial P_i(\theta_a; a^*_{i2}, b^*_{i2}, c_{i2})}{\partial b^*_{i2}} - \frac{a_{i2}}{A^2} \frac{\partial P(\theta_a, a^*_{i2}, b^*_{i2}, c_{i2})}{\partial a^*_{i2}} \right) . \quad (10)$$

Also,

$$\frac{\partial \hat{\xi}^*_a}{\partial B} \equiv \sum_{i=1}^{n} \frac{\partial P_i(\theta_a; a^*_{i2}, b^*_{i2}, c_{i2})}{\partial b^*_{i2}} \frac{\partial b^*_{i2}}{\partial B} . \quad (11)$$

From equation (1), $\frac{b^*_{i2}}{B} = 1$, and substitution into (11) gives

$$\frac{\partial \hat{\xi}^*_a}{\partial B} \equiv \sum_{i=1}^{n} \frac{\partial P_i(\theta_a; a^*_{i2}, b^*_{i2}, c_{i2})}{\partial b^*_{i2}} . \quad (12)$$

The functional form of the partial derivatives of the item response function depends on the mathematical model chosen. Formulas for the partial derivatives for the three-parameter logistic item response function are given in Lord (1980, Chapter 4).

Once the functional form for the item response function is chosen, its derivatives are substituted into equations (10) and (12). These new expressions are then substituted into equations (7) and (8) to find the location of the minimum of F in equation (6).

In the applications described in the following section, the arbitrary group of examinees over which the function was minimized was chosen to be a spaced sample of about 200 examinees from the first calibration of a test. The parameters A and B of the linear transformation were found by minimizing F using the multivariate search technique by Davidon (1959) and Fletcher and Powell (1963).

Results

The Data and Analyses

Data from about 2000 examinees from each of 12 separate administrations of the Scholastic Aptitude Test (SAT) were selected for this study.

The SAT consists of six, 30-minute sections: two operational verbal sections, two operational mathematical sections, one Test of Standard Written English (TSWE) and one variable section containing equating or pretest items. The two verbal sections contain 40 and 45 items respectively; mathematical sections are 25 and 35 items respectively. Verbal equating or pretest sections are 40 items long; corresponding mathematical sections are 25 items long. TSWE data were not used in this study.

Each box in Exhibit 1 represents the operational sections, either verbal or mathematical, of a particular form of the SAT (upper case letters and numbers) and the equating section administered with that test form

(lower case letters). Each box contains items that are the same as items shown in boxes above and below it. For example, the second box in the verbal series contains items designated "X2fe." The "fe" items overlap with those contained in the box labeled "V4fe"; the "X2" items overlap with those contained in the box labeled "X2fm." The last box in each of the verbal and mathematical series contains items that overlap with the items in the first box, thus forming a closed chain.

Each box represents a separate calibration run using the computer program LOGIST (Wingersky, in press; Wingersky, Barton, Lord, 1982). For both the verbal chain and the mathematical chain, the scale established by the calibration of the items in the first box in the chain was arbitrarily chosen as the "base scale" for that chain. The estimates of item parameters for the overlapping items were then used to transform the scales established by the separate calibrations onto the appropriate base scale. For the verbal chain, for example, X2fe was transformed to the scale of V4fe using the item parameter estimates for the fe items that appear in both calibrations. Then X2fm was transformed to the scale of the transformed X2fe items, using the item parameter estimates for the X2 common items. This, of course, places the X2fm items on the V4fe scale. The next set of items, Y3fm, was transformed to the scale of the transformed X2fm items and so forth, until all items were placed on the scale of V4fe.

This sequential transformation process was performed in two ways: (1) the robust mean and sigma Lord and Stocking method described in

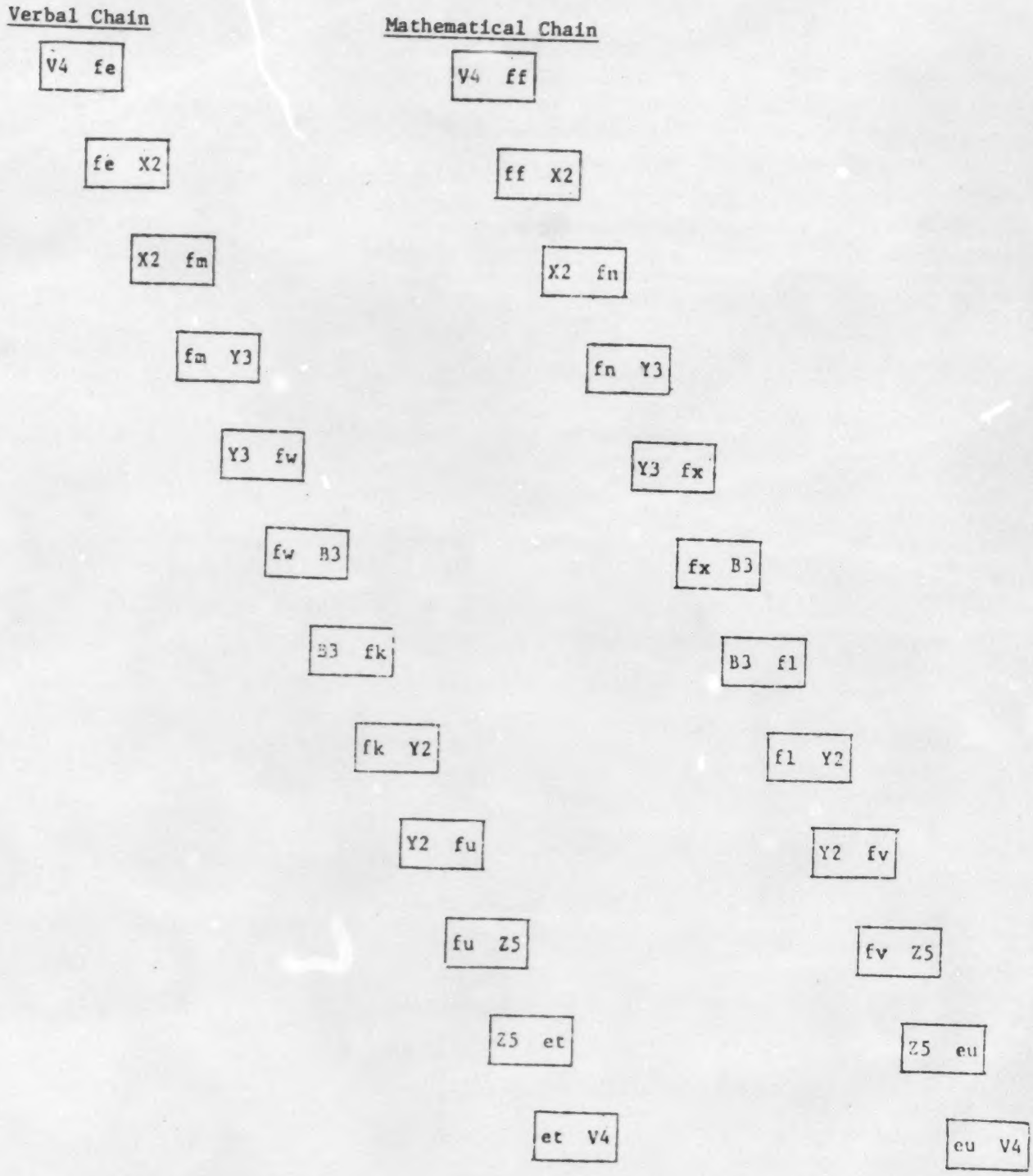

Exhibit 1: Verbal and Mathematical Chains. Each box contains verbal or mathematical sections (capital letters and numbers) and an equating section (small letters).

the Appendix and (2) the new characteristic curve method described previously. This allows the comparison of the end results of the chaining process between the two transformation methods, but does not allow the comparison of the results of individual "links" in the chain.

To compare individual links in the chain, each link in the chain from the robust mean and sigma method was repeated exactly with the characteristic curve method. For example, in the verbal chain, X2fm was transformed to the scale of the (mean and sigma) transformed X2fe by the mean and sigma method as part of the sequential chaining using this method. This link was repeated exactly by using the characteristic curve method to transform X2fm to the scale of the (mean and sigma) transformed X2fe. In contrast to the chain of characteristic curve transformations, this series of characteristic curve transformations does not form a chain.

Results of Transformations for Verbal Items--Individual Links

A typical comparison of individual links is shown in Figures 1 and 2. In Figure 1, the horizontal axis is the (robust mean and sigma) transformed item difficulties for operational section X2 from the X2fe calibration. The vertical axis is the scale of the item difficulties for operational section X2 from the X2fm calibration. In Figure 2, the horizontal axis is the scale of the (robust mean and sigma) transformed item discriminations from X2 of X2fe. The vertical axis is the scale of the item discriminations of X2 from X2fm. The solid line through the

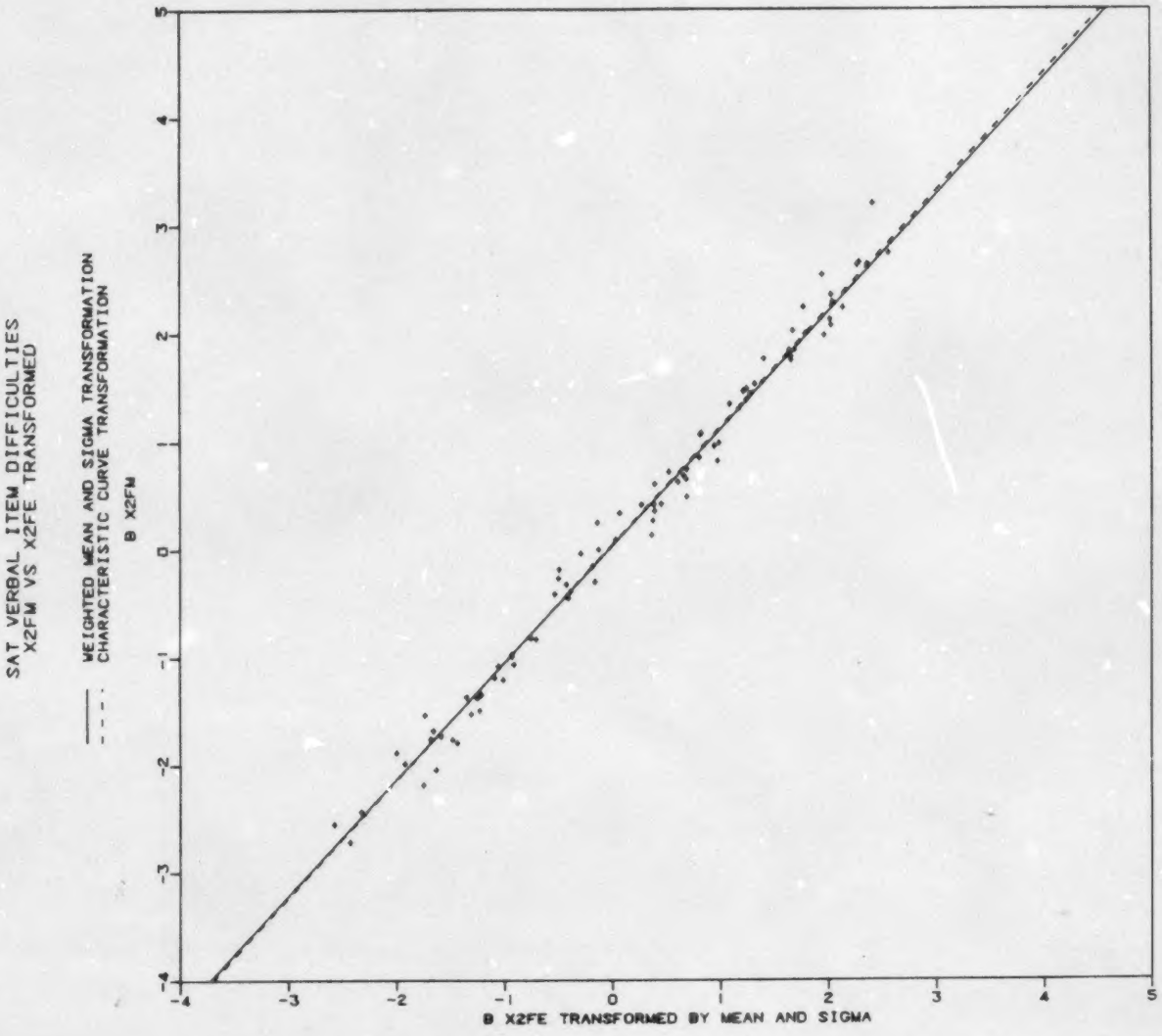

Figure 1. The two transformations for item difficulties compared for a typical verbal link.

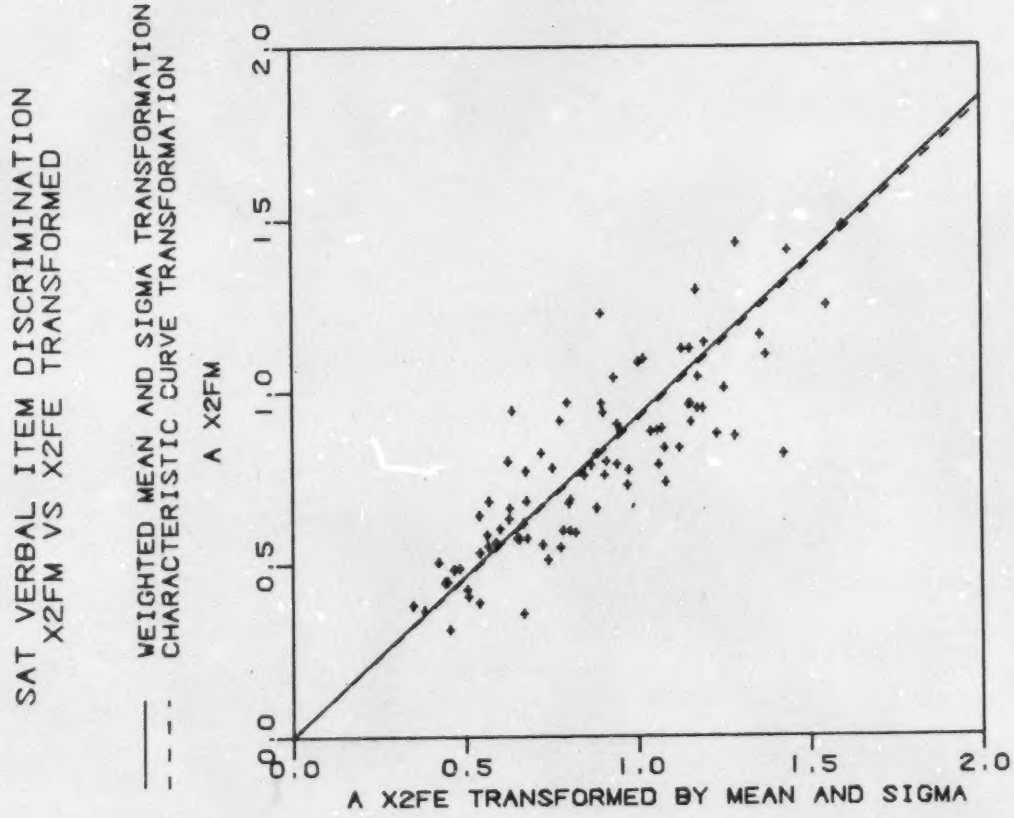

Figure 2. The two transformations for item discriminations compared for a typical verbal link.

the points in each figure is the linear transformation estimated by the robust mean and sigma method. The dashed line is the linear transformation estimated by the new characteristic curve method. The linear transformations do not differ much.

The largest difference found between the two methods for the verbal chain is shown in Figures 3 and 4. Figure 3 shows the presence of six points which could be considered outliers. The robust mean and sigma method explicitly tries to deal with these points, first by giving them low weights if the estimated standard errors are large, and then by giving them low weights if the perpendicular distance to the initial line is large. These points all ended up with weights which were very small or zero, thus some available information may have been discarded. The characteristic curve method does not discard any information. No other verbal link contained as many outliers as this one. It is possible that the difference between the two methods is due to their differential discarding of information.

On the whole, the direct comparison of individual links shows little difference between the two transformation methods for verbal data.

Results of Transformations for Mathematical Items--Individual Links

Most of the comparisons of the two transformation methods using mathematical data show little difference between the two methods. There are exceptions, one of which is shown in Figures 5 and 6. Inspection of Figure 5 shows the characteristic curve transformation is clearly a better fit to the data than the robust mean and sigma

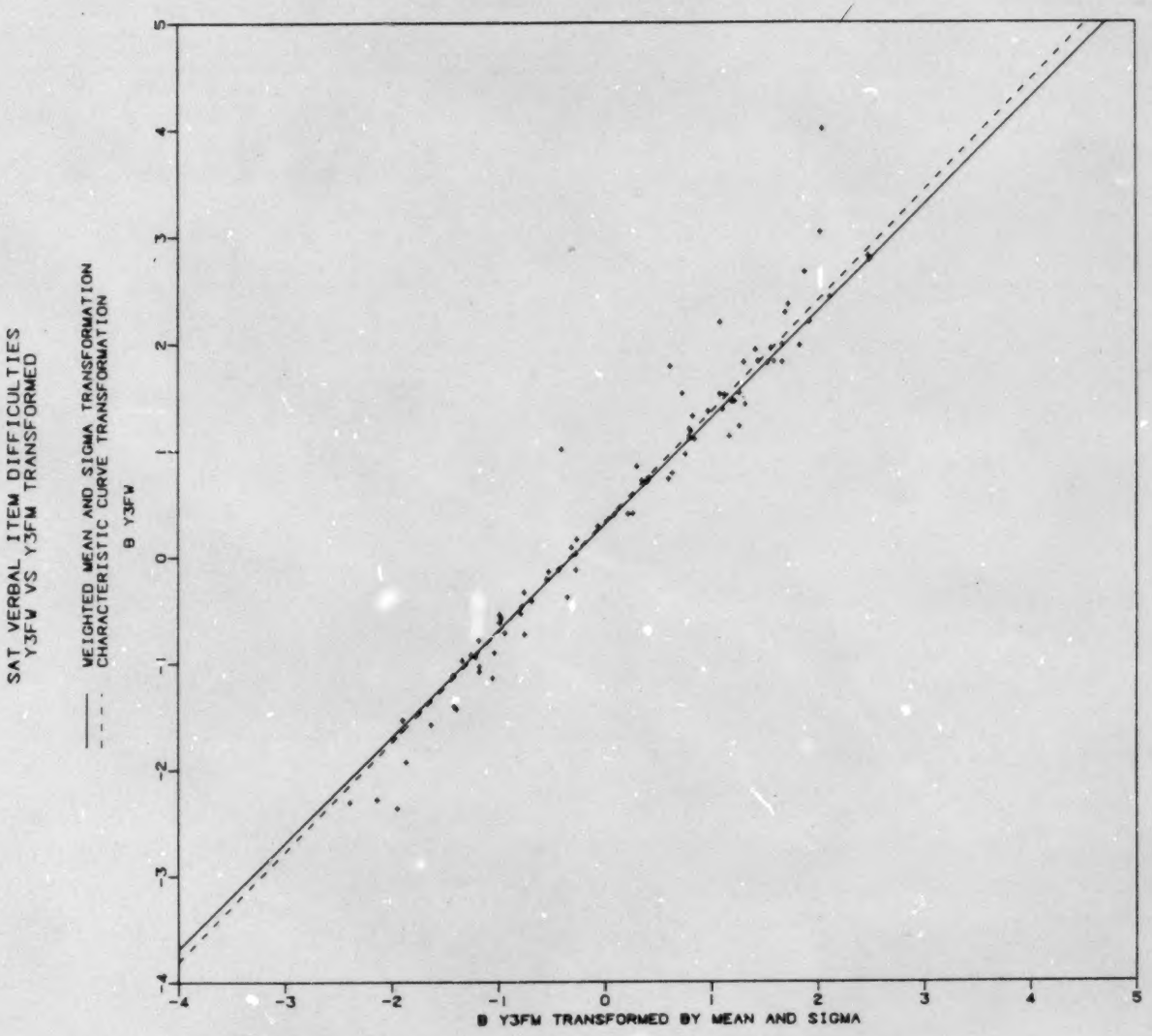

Figure 3. The two transformations for item difficulties compared for the worst verbal link.

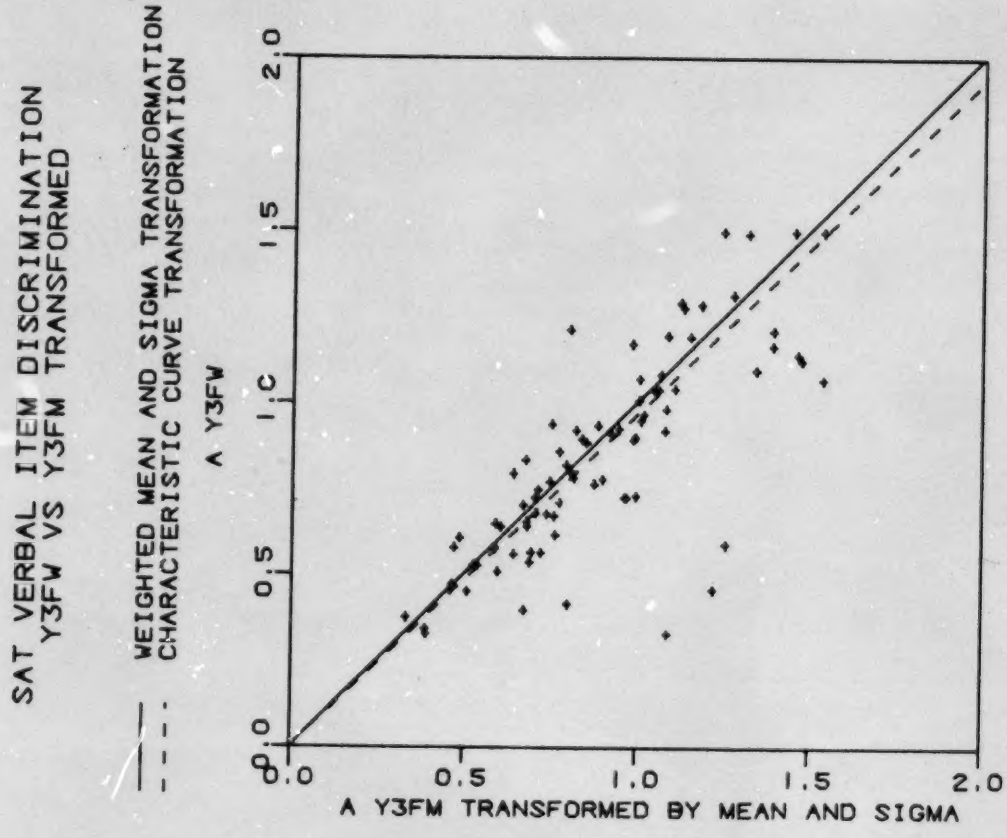

Figure 4. The two transformations for item discriminations compared for the worst verbal link.

transformation. This difference is more visible in Figure 6 where the robust mean and sigma transformation of the item discriminations produces unsatisfactory results. The line does not bisect the point cloud; there are only 18 out of 60 points below the line. The characteristic curve transformation was better; 31 out of 60 points are below the line.

There were two links which produced comparisons of this kind. That is, the characteristic curve transformation worked better than the mean and sigma transformation in both the fit to the item difficulties and the fit to the item discriminations. There were no links in which the mean and sigma transformation fit both the item difficulties and item discriminations better.

Chain Results

The cumulative results of chains of transformations may be evaluated by transforming the last (transformed) set of items in the chain directly to the base scale defined by the first set of items. Since the first and last sets of items are identical, this transformation should be an identity transformation. Figure 7 shows this comparison of each transformation method for the SAT verbal chain, and the identity transformation. The difficulties for items common to the first and last set of items are plotted on the horizontal axis. Figure 8 displays the same information for the SAT mathematical chain.

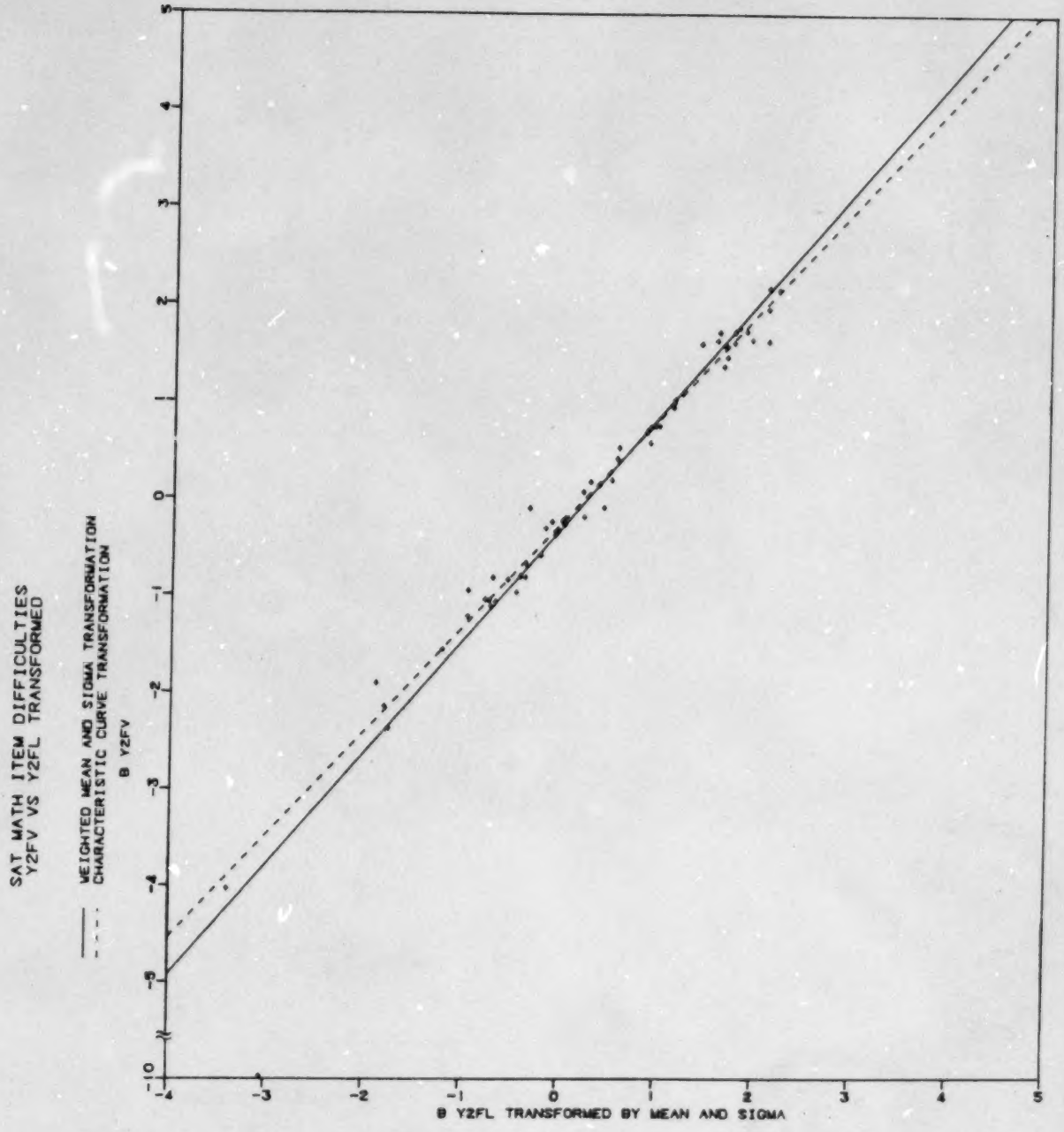

Figure 5. The two transformations for item difficulties compared for a bad mathematical link.

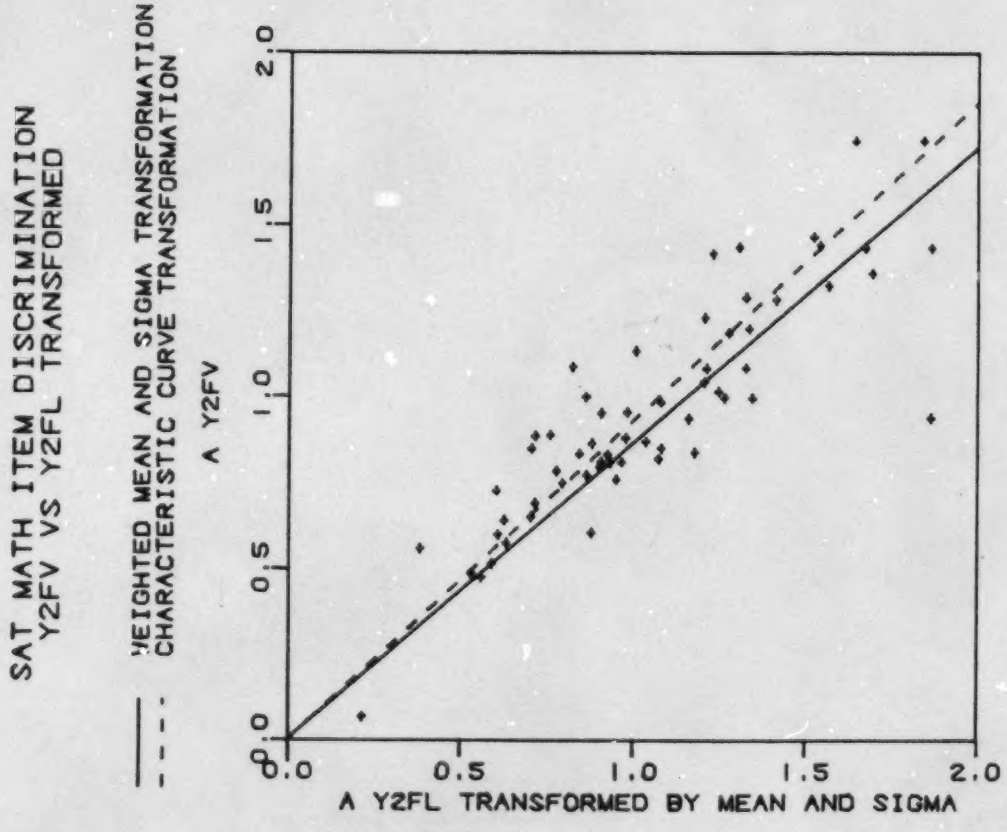

Figure 6. The two transformations for item discriminations compared for a bad mathematical link.

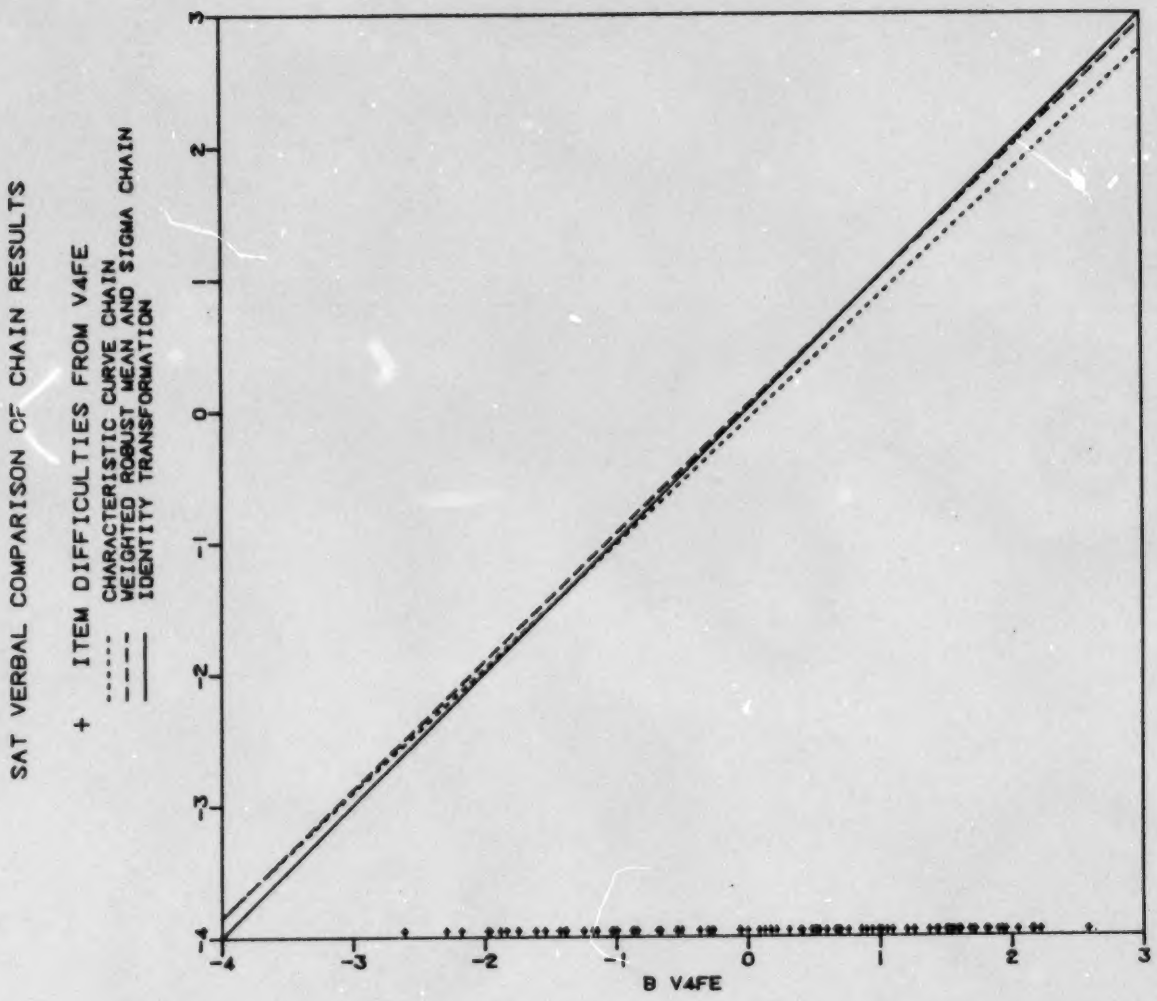

Figure 7. The final transformations for the SAT verbal chain.

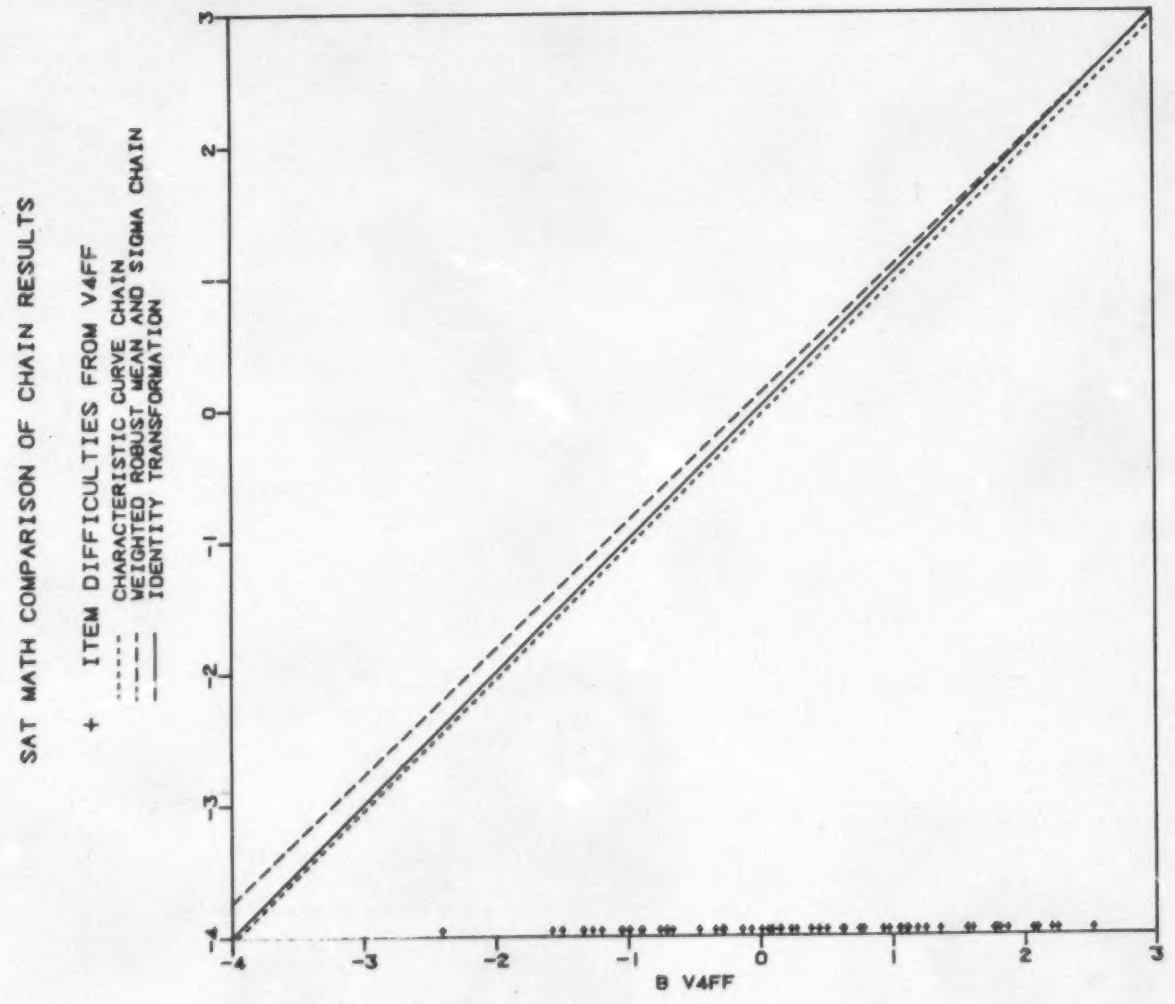

Figure 8. The final transformations for the SAT mathematical chain.

The robust mean and sigma method gives slightly better results than the characteristic curve method for verbal data. For mathematical data, the characteristic curve method worked better than the robust mean and sigma method.

Conclusions

In situations where the robust mean and sigma transformation method worked well, as in the verbal data and most of the mathematical data, the characteristic curve method also worked well. However, the robust mean and sigma method sometimes produced unsatisfactory results. In these instances, the characteristic curve method worked much better. In particular, the characteristic curve method produced a much better transformation for the item discriminations (see Figure 6). If one is choosing a transformation method, the characteristic curve method, which uses more of the information available from each of the calibrations, would be recommended by the authors.

Appendix

Transforming Logistic Scales Using a Robust Iterative Weighted Mean and Sigma Method

This transformation method uses a function of the estimated standard errors of the estimated item difficulties for common items as weights to determine an initial transformation line based on mean and sigma equating of weighted estimates of item difficulties for the common items. A new set of weights is computed using a combination of the estimated standard error weights and robust (Tukey) weights based on perpendicular distances to the line. A new transformation line is computed and the procedure iterates until the maximum change in the perpendicular distances is less than some criterion.

Method

Computing the Standard Errors

The inverse of the information matrix I (p. 191 of Lord (1980)) is an approximation to the variance/covariance matrix for the item parameter estimates. The diagonal element of the inverse corresponding to the item difficulty is the estimated variance of the estimate of item difficulty. The sequare root of this quantity is the estimated standard error of the estimate of item difficulty.

Each item has two estimated item difficulties, one from each calibration. Therefore, each item has two estimated standard errors. The initial weight for an item to be used in the iterative procedure is the reciprocal of the larger estimated squared standard error of the estimated item difficulty.

The accuracy with which an estimated standard error of b is computed is the ratio of the determinant to the product of the diagonals of the information matrix. If this ratio is less than 0.0001, the estimated standard error is not accurate. The item is given a standard error weight of zero.

All people are included in the computation except those who did not reach the item.

Computing the Mean and Sigma Transformation

We have two distributions of weighted estimated item difficulties, one from each calibration. We let $\underset{\sim}{b}_1$ be the distribution from the first calibration, and $\underset{\sim}{b}_2$ be the distribution from the second calibration and compute

$\bar{X}_{\underset{\sim}{b}_1}$, the mean of $\underset{\sim}{b}_1$,

$\sigma_{\underset{\sim}{b}_1}$, the standard deviation of $\underset{\sim}{b}_1$,

$\bar{X}_{\underset{\sim}{b}_2}$, the mean of $\underset{\sim}{b}_2$,

$\sigma_{\underset{\sim}{b}_1}$, the standard deviation of $\underset{\sim}{b}_2$.

The mean and sigma transformation (line) to put the second calibration estimated item difficulties onto the scale of the first is

$$\underset{\sim}{b}_2' = A * \underset{\sim}{b}_2 + B \quad ,$$

where $\underset{\sim}{b}_2'$ is the transformed distribution from the second calibration. For this transformation,

$$A = \sigma_{\underset{\sim}{b}_1} / \sigma_{\underset{\sim}{b}_2} \quad ,$$

$$B = \bar{X}_{\underset{\sim}{b}_1} - A * \bar{X}_{\underset{\sim}{b}_2} \quad .$$

Computing the Tukey Weights

Page 20 of Mosteller and Tukey (1977) gives a method of computing a robust estimate of location by weighting data with differential weights. We use only one piece of this process, namely the formula for the weights.

For our purposes, Y^* is the transformation line we have tentatively found. We replace Tukey's $(Y(i) - Y^*)$ with the perpendicular distance of a point to the line.

Let $D(i)$ equal the absolute value of the perpendicular distance. Then our weights, $T(i)$, are

$$T(i) = \begin{cases} \{1 - (D(i)/CS)^2\}^2 & \text{when } (D(i)/CS)^2 < 1 \\ 0 & \text{otherwise} \end{cases}$$

where S is the median of the $D(i)$ and C is a constant equal to 6.

The Iterative Procedure

The iterative procedure is as follows:

Step 1: For each item difficulty, for each common item, compute

$$W(i) = SE(B(i))^{-2} ,$$

where $SE(B)$ is the larger of the two estimated standard errors.

Step 2: Compute a vector of scaled weights

$$W(i)' = W(i)/(\text{sum of } W(i)) .$$

Step 3: Compute the mean and sigma transformation line between the two sets of estimated item difficulties weighted by W', and get the slope, A, and the intercept, B.

Step 4: Compute the perpendicular distances of each point to the line.

Step 5: Compute the Tukey weights, $T(i)$ for each item, using these perpendicular distances.

Step 6: Reweight each point by a combined weight $U(i)$, where

$$U(i) = (W(i) * T(i))/(\text{sum of } W(i) * T(i)) \quad .$$

Step 7: Compute the weighted mean and sigma transformation line using these new weights.

Step 8: Repeat Steps 4, 5, and 6 until the maximum change in the perpendicular distances is less than 0.01.

Result

This procedure gives low weights to poorly determined item difficulties or to item difficulties which are outliers. Once the final transformation is found for the estimated item difficulties, the estimated item discriminations are transformed, as well as the ability estimates.

References

Bejar, I., & Wingersky, M. S. An application of item response theory to equating the Test of Standard Written English (College Board Report No. 81-8 and Educational Testing Service Research Report 81-35). Princeton, N.J.: College Board Publication Orders, 1981.

Cook, L. L., Eignor, D. R., & Hutton, L. R. Considerations in the application of latent trait theory to objectives-based criterion-referenced tests. Paper presented at the meeting of the American Educational Research Association, San Francisco, 1979.

Davidon, W. C. Variable metric method for minimization (Research and Development Report ANL-5990, rev. ed.). Argonne, Ill.: Argonne National Laboratory, U.S. Atomic Energy Commission, 1959.

Divgi, D. R. Evaluation of scales for multilevel test batteries. Revision of paper presented at the meeting of the American Educational Research Association, Boston, 1980.

Fletcher, R., & Powell, M. J. D. A rapidly convergent descent method for minimization. The Computer Journal, 1963, 6, 163-168.

Haebara, T. Equating logistic ability scales by a weighted least squares method. Japanese Psychological Research, 1980, 22, 144-149.

Linn, R. L., Levine, M. V., Hastings, C. N., & Wardrop, J. L. An investigation of item bias in a test of reading comprehension (Technical Report No. 163). Urbana, Ill.: Center for the Study of Reading, University of Illinois, 1980.

Lord, F. M. Applications of item response theory to practical testing problems. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Marco, G. L. Item characteristic curve solutions to three intractable testing problems. Journal of Educational Measurement, 1977, 14, 139-160.

Mosteller, F. & Tukey, J. W. Data analysis and regression: A second course in statistics. Reading, Mass.: Addison-Wesley, 1977.

Petersen, N. S., Cook, L. L., & Stocking, M. L. IRT versus conventional equating methods: A comparative study of scale stability. Paper presented at the meeting of the American Educational Research Association, Los Angeles, 1981.

Wingersky, M. S. LOGIST: A program for computing maximum likelihood procedures for logistic test models. In R. K. Hambleton (Ed.), ERIBC monograph on applications of item response theory. Vancouver, B.C.: Educational Research Institute of British Columbia, in press.

Wingersky, M. S., Barton, M. A., & Lord, F. M. LOGIST user's guide. Princeton, N.J.: Educational Testing Service, 1982.

DISTRIBUTION LIST

Navy

1 Dr. Jack R. Borsting
Provost and Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Chief of Naval Education and Training Liason Office
Air Force Human Resource Laboratory
Flying Training Division
Williams Air Force Base, AZ 85224

1 CDR Mike Curran
Office of Naval Research
800 North Quincy Street
Code 270
Arlington, VA 22217

1 Dr. Pat Federico
Navy Personnel R & D Center
San Diego, CA 92152

1 Mr. Paul Foley
Navy Personnel R & D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R & D Center
San Diego, CA 92152

1 Patrick R. Harrison
Psychology Course Director
Leadership and Law Department (7b)
Division of Professional Development
U.S. Naval Academy
Annapolis, MD 21402

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. William L. Maloy
Principal Civilian Advisor for Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 CAPT Richard L. Martin, USN
Prospective Commanding Officer
USS Carl Vinson (CNV-70)
Newport News Shipbuilding and Drydock Co.
Newport News, VA 23607

1 Dr. James McBride
Navy Personnel R & D Center
San Diego, CA 92152

1 Mr. William Nordbrock
Instructional Program Development
Building 90
NET-PDCD
Great Lakes NTC, IL 60088

1 Library, Code P201L
Navy Personnel R & D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Psychologist
ONR Branch Office
Building 114, Section D
666 Summer Street
Boston, MA 02210

1 Office of Naval Research
Code 437
800 North Quincy Street
Arlington, VA 22217

5 Personnel and Training Research Programs
Code 458
Office of Naval Research
Arlington, VA 22217

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Office of the Chief of Naval Operations
Research Development and Studies Branch
OP-115
Washington, DC 20350

1 The Principal Deputy Assistant Secretary of the Navy (MRA&L)
4E780, The Pentagon
Washington, DC 22203

1 Director, Research and Analysis Division
Plans and Policy Department
Navy Recruiting Command
4015 Wilson Boulevard
Arlington, VA 22203

1 Mr. Arnold Rubenstein
Office of Naval Technology
800 N. Quincy Street
Arlington, VA 22217

1 Dr. Worth Scanland, Director
Research, Development, Test and Evaluation
N-5
Naval Education and Training Command
NAS
Pensacola, FL 32508

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Alfred F. Smode
Training Analysis and Evaluation Group
Department of the Navy
Orlando, FL 32813

1 Dr. Richard Sorensen
Navy Personnel R & D Center
San Diego, CA 92152

1 Mr. J. B. Sympson
Naval Personnel R & D Center
San Diego, CA 92152

1 Dr. Ronald Weitzman
Code 54 WZ
Department of Administrative Services
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Wisher
Code 309
Navy Personnel R & D Center
San Diego, CA 92152

1 Dr. Martin F. Wiskoff
Navy Personnel R & D Center
San Diego, CA 92152

1 Mr. John H. Wolfe
Code P310
U.S. Navy Personnel Research
and Development Center
San Diego, CA 92152

1 Mr. Ted M. I. Yellen
Technical Information Office
Code 201
Navy Personnel R & D Center
San Diego, CA 92152

Army

1 Technical Director
U.S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Michael Kaplan
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Milton S. Katz
Training Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Harold F. O'Neil, Jr.
Attn: PERI-OK
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 LTC Michael Plummer
Chief, Leadership and Organizational
Effectiveness Division
Office of the Deputy Chief of Staff
for Personnel
Department of the Army
The Pentagon
Washington, DC 20301

1 Dr. James L. Raney
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Mr. Robert Ross
U.S. Army Research Institute for
the Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Sasmor
U.S. Army Research Institute for
the Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Commandant
U.S. Army Institute of Administration
Attn: Dr. Sherrill
Ft. Benjamin Harrison, IN 46256

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 Air Force Human Resources Laboratory
AFHRL/MPD
Brooks Air Force Base, TX 78235

1 U.S. Air Force Office of
Scientific Research
Life Sciences Directorate
Bolling Air Force Base
Washington, DC 20332

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks Air Force Base, TX 78235

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling Air Force Base
Washington, DC 20332

1 Dr. David R. Hunter
AFHRL/MOAM
Brooks Air Force Base, TX 78235

1 Research and Measurement Division
Research Branch, AFMPC/MPCYPR
Randolph Air Force Base, TX 78148

1 Dr. Malcolm Ree
AFHRL/MP
Brooks Air Force Base, TX 78235

Marines

1 Dr. H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Director, Office of Manpower
Utilization
HQ, Marine Corps (MPU)
BCB, Building 2009
Quantico, VA 22134

1 Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

1 MAJ Michael L. Patrow, USMC
Headquarters, Marine Corps
Code MPI-20
Washington, DC 20380

1 Dr. A. L. Slafkosky
Scientific Advisor
Code RD-1
HQ, U.S. Marine Corps
Washington, DC 20380

Coast Guard

1 Chief, Psychological Research Branch
U.S. Coast Guard (G-P-1/2/TP42)
Washington, DC 20593

1 Mr. Thomas A. Warm
U.S. Coast Guard Institute
P.O. Substation 18
Oklahoma City, OK 73169

Other DoD

1 DARPA
1400 Wilson Boulevard
Arlington, VA 22209

12 Defense Technical Information Center
Cameron Station, Building 5
Attn: TC
Alexandria, VA 22314

1 Dr. William Graham
Testing Directorate
MEPCOM/MEPCT-P
Ft. Sheridan, IL 60037

1 Director, Research and Data
OASD (MRA&L)
3B919, The Pentagon
Washington, DC 20301

1 Military Assistant for Training
and Personnel Technology
Office of the Under Secretary of
Defense for Research and Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 Dr. Wayne Sellman
Office of the Assistant Secretary
of Defense (MRA&L)
2B269 The Pentagon
Washington, DC 20301

Civil Government

1 Mr. Richard McKillip
Personnel R & D Center
Office of Personnel Management
1900 E Street, NW
Washington, DC 20415

1 Dr. Andrew R. Molnar
Science Education Development
and Research
National Science Foundation
Washington, DC 20550

1 Dr. H. Wallace Sinaiko
Program Director
Manpower Research and
Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

1 Dr. Vern W. Urry
Personnel R & D Center
Office of Personnel Management
1900 E Street, NW
Washington, DC 20415

1 Dr. Joseph L. Young, Director
Memory and Cognitive Processes
National Science Foundation
Washington, DC 20550

Non-Government

1 Dr. James Algina
University of Florida
Gainesville, FL 32611

1 Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

1 Psychological Research Unit
Department of Defense (Army Office)
Campbell Park Offices
Canberra, ACT 2600
AUSTRALIA

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08541

1 CAPT J. Jean Belanger
Training Development Division
Canadian Forces Training System
CFTSHQ, CFB Trenton
Astra, Ontario K0K 1B0
CANADA

1 Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramat Aviv 69978
ISRAEL

1 Dr. Werner Birke
DezWPs im Streitkraefteamt
Postfach 20 50 3
D-5300 Bonn 2
WEST GERMANY

1 Dr. R. Darrell Bock
Department of Education
University of Chicago
Chicago, IL 60637

1 Liaison Scientists
Office of Naval Research
Branch Office, London
Box 39
FPO, NY 09510

1 Dr. Robert Brennan
American College Testing Programs
P.O. Box 168
Iowa City, IA 52240

1 Dr. C. Victor Bunderson
WICAT Inc.
University Plaza, Suite 10
1160 S. State Street
Orem, UT 84057

1 Dr. John B. Carroll
Psychometric Laboratory
University of North Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. Kenneth E. Clark
College of Arts and Sciences
University of Rochester
River Compus Station
Rochester, NY 14627

1 Dr. Norman Cliff
Department of Psychology
University of Southern California
University Park
Los Angeles, CA 90007

1 Dr. William E. Coffman
Director, Iowa Testing Programs
334 Lindquist Center
University of Iowa
Iowa City, IA 52242

1 Dr. Meredith P. Crawford
American Psychological Association
1200 17th Street, N
Washington, DC 20036

1 Dr. Fritz Drasgow
Yale School of Organization and
Management
Yale University
Box 1A
New Haven, CT 06520

1 Dr. Mike Durmeyer
Instructional Program Development
Building 90
NET-PDCD
Great Lakes NTC, IL 60088

1 ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. A. J. Eschenbrenner
Dept. E422, Bldg. 81
McDonald Douglas Astronautics Co.
P.O. Box 516
St. Louis, MO 63166

1 Dr. Benjamin A. Fairbank, Jr.
McFann-Gray and Associates, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

1 Dr. Leonard Feldt
Lindquist Center for Measurement
University of Iowa
Iowa City, IA 52242

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Dr. Victor Fields
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Univ. Prof. Dr. Gerhard Fischer
Psychologisches Institut der
Universitat Wien
Liebiggasse 5/3
A 1010 Wien
AUSTRIA

1 Prof. Donald Fitzgerald
University of New England
Armidale, New South Wales 2351
AUSTRALIA

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organization
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt, Beranek, and Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Robert Glaser
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. Daniel Gopher
Industrial and Management Engineering
Technion-Israel Institute of
Technology
Haifa
ISRAEL

1 Dr. Bert Green
Department of Psychology
Johns Hopkins University
Charles and 34th Streets
Baltimore, MD 21218

1 Dr. Ron Hambleton
School of Education
University of Massachusetts
Amherst, MA 01002

1 Dr. Delwyn Harnisch
University of Illinois
242b Education
Urbana, IL 61801

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 93106

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Jack Hunter
2122 Coolidge Street
Lansing, MI 48906

1 Dr. Huynh Huynh
College of Education
University of South Carolina
Columbia, SC 29208

1 Prof. John A. Keats
Department of Psychology
University of Newcastle
Newcastle, New South Wales 2308
AUSTRALIA

1 Mr. Jeff Kelety
Department of Instructional Technology
University of Southern California
Los Angeles, CA 90007

1 Dr. Michael Levine
Department of Educational Psychology
210 Education Building
University of Illinois
Champaign, IL 61801

1 Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
NETHERLANDS

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Dr. James Lumsden
Department of Psychology
University of Western Australia
Nedlands, Western Australia 6009
AUSTRALIA

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08541

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77004

1 Dr. Samuel T. Mayo
Loyola University of Chicago
820 North Michigan Avenue
Chicago, IL 60611

1 Prof. Jason Millman
Department of Education
Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Melvin R. Novick
356 Lindquist Center for Measurement
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army Navy Drive
Arlington, VA 22202

1 Dr. Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 Mr. Luigi Petrullo
2431 North Edgewood Street
Arlington, VA 22207

1 Dr. Diane M. Ramsey-Klee
R-K Research and System Design
3947 Ridgemont Drive
Malibu, CA 90265

1 Mr. Minrat M. L. Rauch
P II 4
Bundesministerium der Verteidigung
Postfach 1328
D-53 Bonn 1
GERMANY

1 Dr. Mark D. Reckase
Educational Psychology Department
University of Missouri-Columbia
4 Hill Hall
Columbia, MO 65211

1 Dr. Andrew Rose
American Institutes for Research
1055 Thomas Jefferson St., NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. Lawrence Rudner
403 Elm Avenue
Takoma Park, MD 20012

1 Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

1 Prof. Fumiko Samejima
Department of Psychology
University of Tennessee
Knoxville, TN 37916

1 Dr. Kazuo Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 980
JAPAN

1 Dr. Edwin Shirkey
Department of Psychology
University of Central Florida
Orlando, FL 32816

1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305

1 Dr. Robert Sternberg
Department of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 Dr. Patrick Suppes
Institute for Mathematical Studies in the Social Sciences
Stanford University
Stanford, CA 94305

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massacuusetts
Amherst, MA 01003

1 Dr. Kikumi Tatsuoka
Computer Based Education Research
Laboratory
252 Engineering Research Laboratory
University of Illinois
Urbana, IL 61801

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 Dr. Robert Tsutakawa
Department of Statistics
University of Missouri
Columbia, MO 65201

1 Dr. Howard Wainer
Educational Testing Service
Princeton, NJ 08541

1 Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 East River Road
Minneapolis, MN 55455

1 Dr. Susan E. Whitely
Psychology Department
University of Kansas
Lawrence, KS 66044

1 Dr. Wolfgang Wildgrube
Streitkraefteamt
Box 20 50 03
D-5300 Bonn 2
WEST GERMANY